
10. NUMERIEKE METHODEN

S.M. Hennekens, E. van der Maarel & A.H.F. Stortelder

10.1 Inleiding

Gedurende de laatste decennia zijn vele computerprogramma's ontwikkeld voor het verwerken van vegetatieopnamen. Het gaat hierbij vooral om numerieke methoden en meer in het bijzonder om multivariate analyse-technieken. Deze technieken worden toegepast wanneer van een reeks van objecten tegelijkertijd verschillende kenmerken worden geanalyseerd. Bij het verwerken van opnamen betreffen deze kenmerken de soorten en hun abundantie/bedekking. Multivariate analyses worden thans vrijwel uitsluitend met computers uitgevoerd en daarom wordt ook wel gesproken van computertechnieken. In de plantensociologie worden deze methoden gebruikt in de synthetische fase van het onderzoek (zie hfst. 6), voor het opstellen van vegetatie-eenheden; daarnaast worden ze ook toegepast voor het opsporen van betrekkingen tussen soorten, opnamen en opnamegroepen enerzijds en milieu-gegevens of historische factoren anderzijds.

De computermatige aanpak maakt het mogelijk om snel inzicht te verkrijgen in de variatie binnen de gegevens en deze te visualiseren; de gegevens worden op een systematische wijze opgeslagen en kunnen gemakkelijk opgevraagd, aangevuld en/of opnieuw bewerkt worden. De verschillende stappen in de verwerkingsprocedure zijn bovendien reproduceerbaar. Een opnamentabel die met de hand is vervaardigd, waarin opnamen zijn samengevoegd op grond van de visuele indruk van gelijkheid tegenover ongelijkheid, kan onmogelijk exact worden gereproduceerd door een andere onderzoeker. Een extra voordeel van de computermatige aanpak is dat vegetatietypen kunnen worden 'ontdekt' die tot dusverre aan het oog ontsnapten waren. Een voorbeeld is de numerieke studie van Krahulec et al. (1986), die leidde tot het onderscheiden van vier associaties in de vegetatie van het uitgestrekte kalksteenplateau ('alvar') van het Zweedse eiland Öland. In elk van de vier associaties bleken endemen bij de kensoorten te horen. De endemen (o.a. *Allium schoenoprasum* var. *alvarense* en *Festuca oelandica*)

waren wel bekend, de alvarvegetatie was goed bekend, maar de associaties kwamen pas naar voren toen het opnamemateriaal numeriek werd bewerkt.

Numerieke methoden zijn ruim veertig jaar geleden in de vegetatiekunde geïntroduceerd. Van der Maarel (1979a), die de geschiedenis van de numerieke classificatie ten behoeve van de plantensociologie tot dan toe samenvat, noemt het werk van Sørensen (1948) als beginpunt. Deze Deense onderzoeker paste een similariteitsindex toe voor de berekening van de floristische overeenkomst tussen opnamen en opnamengroepen en stelde op basis hiervan een hiërarchisch classificatiesysteem op. Vervolgens vergeleek hij zijn resultaten met het systeem van Braun-Blanquet.

Met de opkomst van de computer werden numerieke methoden op steeds grotere schaal toegepast, waarbij het gaandeweg mogelijk werd om meer opnamen steeds sneller te verwerken. Bovendien werden -en worden- voortdurend nieuwe numerieke methoden ontwikkeld. Het is dan ook nauwelijks mogelijk om het overzicht te behouden. Samenvattende overzichten worden, behalve door Van der Maarel (1979a), onder meer gegeven door Gauch (1982), Goodall (1978), Pielou (1984), Kershaw & Looney (1985), Jongman et al. (1987), Austin (1985), Noy-Meir & Van der Maarel (1987) en Mucina & Van der Maarel (1989). Vooral voor de niet-specialist is het niet eenvoudig om bij een gegeven wetenschappelijke vraagstelling een geschikte methode te kiezen uit de grote scala aan mogelijkheden. De keuze voor een bepaalde methode wordt in publicaties veelal niet gemotiveerd en lijkt sterk beïnvloed te worden door de lokale beschikbaarheid van programma's. Het gevaar bestaat derhalve dat onvoldoende kennis van de gebruikte techniek leidt tot een weinig kritisch of zelfs onjuist gebruik. Er zijn weliswaar diverse pogingen ondernomen om de toepasbaarheid van de afzonderlijke methoden te evalueren (o.a. Everitt 1974; Gauch 1982; Van Tongeren 1987), maar hieruit is nooit een eensluidende opvatting naar voren gekomen (zie ook Kent & Ballard 1988; Mucina & Van der Maarel 1989).



Foto 10.1. Het met de hand samenstellen van plantensociologische tabellen was een tijdrovende zaak; door gebruik te maken van een zogenaamd tabellenbord kon de procedure enigszins worden versneld.

Multivariate methoden worden onderverdeeld in classificatie- en ordinatietechnieken. Zowel bij classificatie als bij ordinatie wordt de plaats die een vegetatieopname krijgt toegewezen bepaald door de floristische samenstelling, waarbij abundantie/bedekkingswaarden al of niet een rol kunnen spelen (Van der Maarel 1979a). Classificatie, in plantensociologische context, kan worden omschreven als de procedure voor het onderscheiden van vegetatietypen; het gaat hierbij om het vaststellen van zo homogeen mogelijke groepen van opnamen ('clusters') die onderling zoveel mogelijk verschillen. Ordinatie is in dit verband het rangschikken van opnamen en/of soorten in een één- of meerdimensionale ruimte, waarbij overeenkomstige opnamen of soorten dicht bij elkaar, en opnamen of soorten die weinig overeenkomen ver uit elkaar worden geplaatst (Goodall 1954). De hypothese daarbij is dat de rangschikking gecorreleerd is met de variatie in een of meer milieuvariabelen, dat wil zeggen met milieu-gradiënten.

In de beginperiode van het gebruik van de multivariate methoden werden classificatie- en ordinatietechnieken gescheiden van elkaar ontwikkeld, in die zin dat niet getracht werd ze op elkaar aan te laten sluiten. Er was veel discussie over de betekenis en de keuze van beide technieken (Kent & Ballard 1988). Tegenwoordig is men van mening dat beide technieken elkaar zeer goed aanvullen en naast elkaar kunnen worden toegepast.

In dit hoofdstuk wordt allereerst ingegaan op de vraag hoe de informatiewaarde van soorten gekwantificeerd kan

worden (§ 10.2). Hierbij wordt aandacht besteed aan transformatie en standaardisatie van bedekkingswaarden van soorten en aan weging van soorten, omdat deze grote invloed hebben op het resultaat van classificaties en ordinaties. Aansluitend worden in paragraaf 10.3 enkele indices besproken die gebruikt worden voor het bepalen van de similariteit tussen opnamen. In paragraaf 10.4 worden de begrippen ordinatie en classificatie nader uitgewerkt; een onderdeel daarvan is een toelichting op de twee clusterprogramma's TWINSpan en FLEXCLUS. Vervolgens wordt bij wijze van voorbeeld de procedure behandeld voor de verwerking van vegetatieopnamen, zoals toegepast bij de classificatie van de vegetatie van Nederland (§ 10.5). De ontwikkeling van numerieke verwerkingsmethoden binnen de plantensociologie is ten dele voortgekomen uit onvrede met de vermeende vooringenomenheid van plantensociologen bij het samenstellen van vegetatietabellen. De resultaten van de handmatige methode zouden in hoge mate bepaald worden door de persoonlijke opvatting van de onderzoeker; het toepassen van numerieke methoden daarentegen zou onbevooroordeeld zijn. In paragraaf 10.6 wordt hierop ingegaan.

10.2 Transformatie, weging en standaardisatie

Wanneer men bij de numerieke verwerking van vegetatieopnamen gebruik maakt van de abundantie/bedekkings-

waarden van soorten, dient eerst de hiervoor in het veld gehanteerde codering getransformeerd te worden. De codes r, +, 2m, 2a en 2b (van de door Barkman, Doing en Segal in 1964 gemodificeerde schaal van Braun-Blanquet) kunnen namelijk niet als zodanig gebruikt worden voor numerieke verwerking; hetzelfde geldt voor de niet-numerieke codes uit andere schalen (zie §5.3). Meestal worden de in het veld gehanteerde opnameschalen omgezet in een negendelige, zogenaamde ordinale schaal (Van der Maarel 1979b; zie ook fig. 5.4 en 5.5). Voor de veel gebruikte gemodificeerde schaal van Braun-Blanquet heeft dit voor de klassenindeling geen consequenties en komt de omzetting neer op een rechtstreekse vertaling van codes in cijfers.

Welk type transformatie het best toegepast kan worden, en hoe daar verder in de procedure mee gewerkt wordt, hangt af van de vraagstelling en de aard van het opname materiaal. Bij het rekenen met bedekkingsverschillen zijn twee uitersten te onderscheiden. In het ene uiterste wordt uitgegaan van de bedekkingen in procenten, waarbij een soort die bijvoorbeeld een bedekking van 37 % heeft ook 37 maal zo zwaar weegt als een soort met een bedekking van 1 %. De soorten met een hoge bedekking zijn hierbij bepalend voor de indeling, terwijl de soorten met een relatief lage bedekking nauwelijks van invloed zijn. In het andere uiterste spelen de bedekkingswaarden géén rol en wordt slechts uitgegaan van de presentie of absentie der soorten. In het algemeen leidt een tussenweg tot de beste resultaten. Het beklemtonen van de presentie of absentie van soorten geeft meestal goede resultaten bij de verwerking van een heterogene reeks opnamen en/of van soortenrijke opnamen. Bij een homogene reeks opnamen zijn het vooral de verschillen in bedekkingswaarden die aanleiding vormen tot het onderscheiden van opnamegroepen. Ook bij soortenarme opnamen, zoals vaak het geval is bij dominantie-gemeenschappen, is een accentuering van de bedekking zinvol. In welke mate de bedekking beklemtoond dient te worden, hangt verder af van de rang van de betrokken syntaxa. Jensén (1978) en Van der Maarel (1979b) stellen dat hoog in de hiërarchie van het classificatiesysteem bedekkingen van soorten een ondergeschikte rol spelen; naarmate men in de hiërarchie afdaalt, worden de bedekkingswaarden steeds meer van belang. Het beperkt houden van de betekenis van hoge bedekkingen van soorten is ook op zijn plaats, omdat dominantie vaak op storing wijst en bij het zwaarder laten wegen van de dominanten de informatie van de nog met lage bedekking aanwezige oorspronkelijke soorten onderbelicht blijft.

Het rekenen met de bovengenoemde negendelige schaal komt goed overeen met de opvatting van Braun-Blanquet dat bij classificatie relatief veel waarde moet worden toegekend aan verschillen in abundantie wanneer de

bedekkingen laag zijn. In het bijzonder de karakteristieke soorten immers komen in veel gevallen met een lage bedekking in de vegetatie voor en juist dan spelen verschillen in abundantie een belangrijke rol. Wanneer reële bedekkingswaarden beschikbaar zijn, wordt om dezelfde reden vaak een logaritmische of een worteltransformatie toegepast, die in wezen overeenkomt met de ordinale transformatie (Van der Maarel 1979b; Van Tongeren 1987).

In de meeste multivariate analysetechnieken wordt de mogelijkheid geboden om op verschillende manieren te rekenen met de bedekkingswaarden der soorten. In het programma TWINSpan bijvoorbeeld kan een soort vervangen worden door één of meer 'pseudosoorten'; naarmate een soort een hogere bedekking heeft, neemt het aantal pseudosoorten toe, waardoor de soort meer gewicht krijgt (Hill et al. 1975). Door het definiëren van drempelwaarden (*cut levels*) kan men het aantal pseudosoorten zelf bepalen. Stel dat, uitgaande van de ordinale schaal (1 t/m 9), de volgende drempelwaarden worden gedefinieerd: 1, 4, 6 en 9, en dat *Poa pratensis* in opname 1 met waarde 8 voorkomt en in opname 2 met waarde 5. De twee opnamen bevatten dan de volgende pseudosoorten:

Opname 1:	Poapra ¹ , Poapra ² , Poapra ³
Opname 2:	Poapra ¹ , Poapra ²

De opnamen komen dus voor twee pseudosoorten met elkaar overeen en verschillen in één pseudosoort. Men kan desgewenst de negendelige ordinale schaal volledig tot zijn recht laten komen door negen drempelwaarden te definiëren.

Behalve door middel van transformatie kan de classificatie beïnvloed (geoptimaliseerd) worden door aan soorten een verschillend gewicht toe te kennen, samenhangend met verschillen in hun diagnostische waarde. In hoofdstuk 7 is aan dit aspect uitvoerig aandacht besteed bij de syntaxonomische identificatie van opnamen en tabellen. Als men bijvoorbeeld bosopnamen wil indelen op basis van de spontane soorten, dienen de soorten in de boomlaag ondergewaardeerd te worden, omdat deze (in Nederland) vaak aangeplant of door selectieve kap sterk beïnvloed zijn. Een ander geval waarbij het toekennen van hogere waarderingen aan sommige soorten verantwoord is, doet zich voor wanneer de classificatie (noodgedwongen) uitgevoerd wordt op basis van heterogeen opname-materiaal, bijvoorbeeld bij het analyseren van historische opnamen, waarbij werd uitgegaan van - naar de huidige opvattingen - te grote proefvlakken. Bij het classificeren van plantengemeenschappen van het open water kan het om deze reden zinvol zijn aan de echte waterplanten een

groter gewicht toe te kennen dan aan de helofyten. Wanneer uitgegaan zou zijn van kleinere proefvlakken, zou het aandeel van deze oeverplanten veel geringer zijn geweest (zie Schaminée et al. 1990).

Standaardisatie is een derde mogelijkheid om een groep van opnamen te bewerken alvorens de clusterprocedure uit te voeren. Hierbij wordt per soort de bedekkingswaarde (of de presentie) gerelateerd aan de som van de waarden in alle opnamen. Deze standaardprocedure in de statistiek is ingebouwd in sommige multivariate technieken en bij gebruik van dergelijke technieken standaardiseert men automatisch. Vanuit ecologisch oogpunt aantrekkelijker is standaardisatie naar de hoogste bedekking die een soort in het materiaal vertoont. Hierbij worden voor iedere soort hetzij de bedekking hetzij de presentie, of beide, gerelateerd aan de hoogste bedekkingswaarde. Standaardisatie naar de maximale bedekking van soorten bijvoorbeeld wordt toegepast wanneer men van mening is dat kleine verschillen in bedekking bij soorten die nooit met hoge bedekkingen voorkomen ecologisch even indicatief zijn als grote bedekkingsverschillen bij dominante soorten. Bij clustering van bosopnamen bijvoorbeeld kan door standaardisatie van de bedekkingswaarden de betekenis van een weinig bedekkende soort uit de kruidlaag gelijk worden gesteld aan die van een dominante boomsoort. Bij de standaardisatie van de presentie van soorten worden zeldzame soorten opgewaardeerd en wordt het gewicht van algemene soorten verminderd (Jongman et al. 1987).

Noest et al. (1989) stellen dat in de meeste methoden die gebruikt worden bij multivariate analyse van vegetatiegegevens te veel nadruk wordt gelegd op enerzijds soorten met een brede ecologische amplitudo, die vaak ook met hoge bedekkingen voorkomen (algemene soorten), of anderzijds op soorten met juist een smalle ecologische amplitudo (zeldzame soorten). Deze bezwaren kunnen tot op zekere hoogte worden ondervangen door transformatie en standaardisatie, zoals hierboven genoemd, maar hieraan blijven toch bezwaren verbonden. Transformatie leidt tot informatieverlies en standaardisatie kan gepaard gaan met een overwaardering van soorten met een lage bedekking. Genoemde auteurs stellen de zogenaamde optimum-transformatie voor, waarbij lage bedekkingswaarden van soorten alleen dan een hoger gewicht krijgen als blijkt dat deze lage waarden ook onder optimale omstandigheden voor de desbetreffende soorten normaal zijn; men gaat dus uit van de optimumrespons van de soorten. Wanneer in een voedselrijk hellingbos *Urtica dioica* (Grote brandnetel) en *Orchis purpurea* (Purperorchis) beide met een bedekking van 5 % voorkomen, wordt bij de optimumtransformatie aan deze soorten een verschillend gewicht toegekend. Omdat *Urtica dioica*

vaak met hoge bedekkingen voorkomt, is een bedekking van 5 % ecologisch onbelangrijk, terwijl een bedekking van 5 % van *Orchis purpurea* een grote betekenis heeft.

10.3 Indices

Het berekenen van de mate van overeenkomst (similariteit) of verschil (dissimilariteit) tussen opnamen vindt plaats met behulp van bepaalde formules, zogenaamde similariteitsindices. In de loop van de tijd is een groot aantal indices ontwikkeld. Aan de basis van alle similariteitsindices ligt het idee ten grondslag dat de overeenkomst groter is naarmate twee opnamen meer soorten gemeenschappelijk hebben en in minder soorten van elkaar verschillen. Een van de eerste indices was die van Jaccard (1902). Hoewel deze index oorspronkelijk was bedoeld om flora's van verschillende gebieden met elkaar te vergelijken, werd hij later ook toegepast om de overeenkomst tussen vegetatie-eenheden te bepalen (zie Barkman 1958). De Jaccard-similariteitsindex (SJ) is als volgt gedefinieerd:

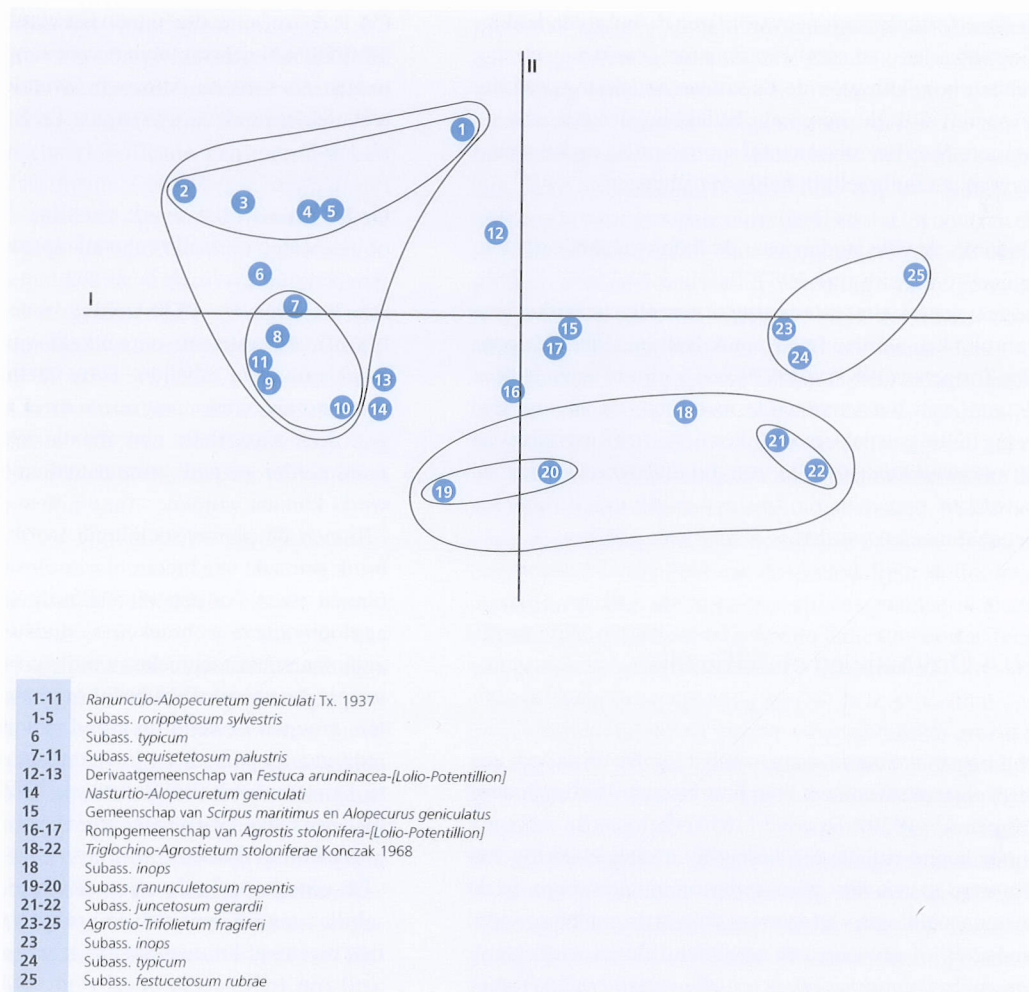
$$SJ = c / (a + b + c) \quad (10.1)$$

waarbij c het aantal gemeenschappelijke soorten weergeeft en a en b betrekking hebben op het aantal soorten dat voor ieder van de opnamen uniek is.

Sørensen stelde in 1948 een andere similariteitsindex voor, waarbij het aantal gemeenschappelijke soorten van beide opnamen niet wordt gedeeld door het totaal aantal soorten in beide opnamen, maar door het gemiddeld aantal soorten. De rol van de gemeenschappelijke soorten heeft hier dus meer nadruk dan bij de Jaccard-index. De coëfficiënt van Sørensen (SS) wordt ook wel community-coëfficiënt genoemd. In formule:

$$SS = 2c / (a + b + 2c) \quad (10.2)$$

De indices kunnen gebruikt worden voor het vergelijken van opnamen of van clusters. In het laatste geval is de presentie van de soorten een belangrijk gegeven, die echter in geen van beide formules is verdisconteerd. Kulczynski (1928) paste de formule van Jaccard in dit opzicht aan. Ellenberg (1956) breidde de formule verder uit door de bedekkingswaarden van de soorten op te nemen. In het programma FLEXCLUS (zie verderop) wordt de similariteitsratio van Ball (1966) gebruikt. Deze index wordt bij de berekening van de similariteit tussen opname i en opname j voorgesteld door de volgende formule:



Figuur 10.1. Ordinatie van de plantengemeenschappen van het Lolio-Potentillion in Nederland. De eerste as correleert met het zoutgehalte van het grondwater, de tweede as met de diepte van het grondwater tijdens het groeiseizoen. De associaties en subassociaties zijn omcirkeld (naar Sykora 1983).

$$SR_{ij} = \sum_k y_{ki} y_{kj} / \left(\sum_k y_{ki}^2 + \sum_k y_{kj}^2 - \sum_k y_{ki} y_{kj} \right) \quad (10.3)$$

Hierin is y_{ki} de bedekking van de k -de soort in opname i . Bij transformatie naar presentie-absentie, waarbij dus de verschillen in bedekkingswaarde wegvallen, is deze formule identiek aan die van Jaccard. Er zijn tal van andere indices voorgesteld voor het bepalen van de mate van overeenkomst tussen vegetatieopnamen, onder andere door Barkman (1958):

$$SR = (c+1) / \sqrt{(ab+1)} \quad (10.4)$$

De toevoeging +1 in teller en noemer is opgenomen, omdat de noemer niet gelijk aan nul mag zijn; het karakter van de index wordt door deze toevoeging nauwelijks

beïnvloed, behalve bij soortenarme opnamen. Een index waarbij niet alleen rekening wordt gehouden met verschillen in bedekking van soorten, maar ook met de som van de bedekkingswaarden en het verschil tussen minimum- en maximumwaarden, wordt gegeven door Noest & Van der Maarel (1989). Hierbij wordt niet de similariteit tussen opnamen berekend maar de dissimilariteit (DR = dissimilariteitsratio):

$$DR_{jk} = \left(\sum_i \left(|x_{ij} - x_{ik}| / (x_{ij} + x_{ik}) + |x_{ij} - x_{ik}| / (x_{\max} - x_{\min}) \right) / 2 \right) / (N - Z_{jk}) \quad (10.5)$$

In deze formule vertegenwoordigen x_{ij} en x_{ik} de bedekkingswaarden van soort i in opname j en k ; $x_{\max}$ en $x_{\min}$ hebben betrekking op de maximale bedekkingswaarde, respectievelijk de minimale bedekkingswaarde van de soorten, N op het totaal aantal soorten en Z_{jk} op het aantal soorten dat ontbreekt in beide opnamen.

Ondanks de vele studies naar de toepassingswaarde van indices (o.a. Faith et al. 1987; Belbin & McDonald 1993), bestaat geen overeenstemming over welke index het best gebruikt kan worden (zie Kent & Ballard 1988). Volgens Van Tongeren (1987) wordt de keuze vooral bepaald door de aard van het verzamelde materiaal, de ecologische vraagstelling en de persoonlijke voorkeur en ervaring van de onderzoeker. In feite zou bij iedere reeks van te bewerken opnamen de keuze van de index opnieuw bepaald moeten worden.

10.4 Ordinatatie en classificatie

Het begrip ordinatatie stamt van het Duitse *Ordnung*, een term die in deze context voor het eerst werd gebruikt door Ramenski (1930). Goodall (1954) definieerde ordinatatie in algemene zin als: een ordening of rangschikking van objecten in een één- of meerdimensionale ruimte. In de plantensociologie vertegenwoordigen de punten soorten, opnamen of groepen van opnamen (clustercentroïden). Omdat het niet mogelijk is om alle variatie in één figuur weer te geven, wordt bij ordinatatie gebruik gemaakt van een aantal complementaire diagrammen. Elk ordinatiedigram is een tweedimensionale weergave, een projectie van alle punten in één vlak, met twee haaks op elkaar staande assen. De assen van de afzonderlijke diagrammen 'verklaren' elk een deel van de variatie. Een voorbeeld van een ordinatiedigram wordt gegeven in figuur 10.1.

Er worden twee vormen van ordinatietechnieken onderscheiden: directe en indirecte. Bij indirecte ordinatatie worden in eerste instantie geen milieugegevens gebruikt, maar wordt uitsluitend uitgegaan van de floristische samenstelling, om op indirecte wijze, gebruik makend van de verkregen ordinatiedigrammen, de belangrijkste milieuv variabelen te achterhalen. Bij directe ordinatatie wordt wel gebruik gemaakt van milieugegevens en is de ordening van de punten (opnamen of soorten) gebaseerd op het verband tussen milieufactoren en vegetatiegegevens (zie o.a. Fresco 1972). De meest gebruikte ordinatietechnieken zijn: PCA (*Principal Component Analysis*), CA (*Correspondence Analysis*, ook wel *Reciprocal Averaging*, RA, genoemd) en DCA (*Detrended Correspondence Analysis*), alle voorbeelden van indirecte ordinatietechnieken.

CA is de techniek die binnen het classificatieprogramma TWINSpan gebruikt wordt voor rangschikking van opnamen en soorten. Voor een overzicht van ordinatietechnieken wordt verwezen naar Ter Braak (in Jongman et al. 1987).

Onder classificatie wordt verstaan: het toekennen van objecten (bijvoorbeeld vegetatieopnamen of soorten) aan groepen of klassen, op basis van hun onderlinge gelijkheid. Tot ongeveer 1970 werden in de plantensociologie classificatiesystemen ontwikkeld op basis van met de hand gemaakte tabellen. Deze methode werd met de opkomst van computers steeds meer verlaten en vervangen door numerieke classificatiemethoden, waarmee, zoals eerder gezegd, grote aantallen opnamen snel verwerkt kunnen worden.

Binnen de plantensociologie wordt voornamelijk gebruik gemaakt van hiërarchische clustermethoden, waarbinnen twee vormen te onderscheiden zijn, te weten agglomeratieve technieken en divisieve technieken. Bij agglomeratieve technieken wordt via een reeks van samenvoegingen vanuit afzonderlijke opnamen naar steeds grotere groepen gewerkt op grond van gelijkenis tussen de individuele opnamen en groepen (*bottom up*). Bij divisieve technieken wordt vanuit de totale verzameling van opnamen een reeks splitsingen uitgevoerd naar steeds kleinere groepen (*top down*).

De criteria op basis waarvan overeenkomsten en verschillen tussen opnamen en groepen van opnamen worden berekend kunnen monothetisch of polythetisch van aard zijn. In het eerste geval vindt splitsing of samenvoeging plaats op grond van aan- of aanwezigheid van slechts één soort, in het tweede geval is meer dan één soort bepalend voor de groepering van opnamen. Omdat in de Braun-Blanquet-benadering de classificatie wordt gebaseerd op de totale floristische samenstelling van de opnamen, worden hier vooral polythetische criteria toegepast.

Beide technieken, zowel agglomeratieve als divisieve, hebben voor- en nadelen. Voor divisieve technieken noemt Everitt (1974) onder andere het probleem van het niet kunnen herplaatsen (re-alloceren) van opnamen die in een vroeg stadium onjuist geclassificeerd zijn. Causton (1988) daarentegen stelt dat een divisieve techniek (zoals TWINSpan) ten opzichte van een agglomeratieve techniek het voordeel heeft dat clusters worden onderscheiden vanuit het overzicht van alle opnamen. Dit voordeel geldt vooral als gewerkt wordt met een groot aantal opnamen, bijvoorbeeld als de clustering gericht is op het ontwikkelen van formele classificatiesystemen. Voor het maken van lokale typologieën, waarbij vaak kleine groepen van opnamen bewerkt worden, voldoen agglomeratieve technieken (zoals verwerkt in het programma FLEXCLUS) in sommige gevallen beter, vooral wanneer de onderlinge verschillen

tussen de opnamen klein zijn. Het programma FLEXCLUS wordt om die reden wel in complementaire zin gebruikt om de clusters van de lagere niveaus van grote TWINSPAN-bewerkingen opnieuw te beoordelen.

Behalve hiërarchische classificatie kan men ook een zogenaamde net-classificatie (*reticulate classification*) toepassen. Hierbij wordt voor een bepaald aantal clusters gekozen en vervolgens de allocatie van opnamen in clusters geoptimaliseerd door middel van re-allocatie. Meestal wordt met een agglomeratieve of divisieve strategie begonnen, bijvoorbeeld in het programma TABORD (Van der Maarel et al. 1978) of in FLEXCLUS. Het is ook mogelijk uit te gaan van het resultaat van een ordinatie en de opnamen op basis van hun configuratie in het ordinatiediagram in een aantal groepen te verdelen (*ordination space partitioning*).

10.4.1 Het programma TWINSPAN

Van het grote aanbod van numerieke classificatiemethoden wordt voor plantensociologische doeleinden het programma TWINSPAN het meest gebruikt, zowel buiten als binnen Nederland. TWINSPAN (Hill 1979) is de afkorting van *TWo-way INdicator SPecies ANalysis*. Het resultaat van het programma wordt onder meer in de vorm van een tabel gepresenteerd, waarin de opnamen en de soorten zodanig geordend zijn, dat in de matrix de getransformeerde bedekkingswaarden van de soorten zoveel mogelijk volgens een diagonaal zijn gerangschikt (zie ook Hill et al. 1975).

De eerste stap bij de TWINSPAN-bewerking is de ordinatie van de opnamen aan de hand van alle soorten door middel van *Reciprocal Averaging*. De eerste ordinatie-as geeft de maximale spreiding van de opnamen weer, een ordening die overeenkomt met de belangrijkste floristische gradiënt. Voor iedere opname wordt een score op deze as berekend en vervolgens wordt een (voorlopige) splitsing aangebracht, zodanig dat het floristische verschil tussen de twee verkregen groepen van opnamen maximaal is. De opnamen aan de linkerkant van het punt van splitsing behoren tot de zogenaamde negatieve groep, die aan de rechterkant tot de zogenaamde positieve groep. Vervolgens wordt berekend hoe de soorten zijn verdeeld over de beide groepen. De soorten die een sterke voorkeur hebben voor de ene dan wel voor de andere groep van opnamen, kunnen voor deze clusters als indicatief worden beschouwd. Hill et al. (1975) spreken van *indicator species*.

De aldus verkregen tweedeling van opnamen wordt vervolgens verfijnd door een herberekening uit te voeren op basis van de zogenaamde preferentiescores van de soorten. De hoogste preferentiescore van 1 houdt in dat de

presentie van een soort in de ene groep van opnamen tenminste driemaal zo hoog is als in de andere. Soorten met lage presentiescores worden ondergewaardeerd. Na deze tweede ordinatie volgt nog een derde en laatste berekening, uitsluitend op basis van de indicatieve soorten. Deze ordinatie heeft nauwelijks nog invloed op de indeling in twee groepen, maar heeft vooral tot doel om na te gaan of de verdeling van de opnamen over de beide groepen overeenkomt met de verdeling van de geselecteerde indicatorsoorten. Wanneer blijkt dat dit voor een bepaalde opname niet het geval is, wordt deze als '*misclassified*' aangegeven. Dit hoeft evenwel nog niet te betekenen dat die opname werkelijk verkeerd geïnclassificeerd is. De opname mist weliswaar de indicatieve soorten die TWINSPAN aangeeft, maar hoort op basis van de overige soorten blijkaar toch tot het cluster. Wel is het zaak de indeling van een dergelijke opname nader te beschouwen in het licht van de uiteindelijke, definitieve classificatie. Het aantal *misclassified* opnamen is mede afhankelijk van het aantal gekozen indicatorsoorten (een van de opties in het TWINSPAN-programma).

Na de definitieve splitsing kan de hele procedure opnieuw beginnen met elk van de twee gevormde groepen van opnamen. Het maximale aantal niveaus waarop aldus splitsingen uitgevoerd worden kan tevoren worden aangegeven, evenals de minimale grootte der clusters. Welke tweedelingen op een laag splitsingsniveau nog zinvol zijn is een vraag die bij de interpretatie van de uiteindelijke indeling moet worden beantwoord, al levert het programma hiervoor wel enige indicatie door bij iedere splitsing aan te geven in welke mate de twee clusters floristisch van elkaar verschillen, de zogenaamde eigenwaarde. Het uiteindelijke aantal typen wordt bij interpretatie vastgesteld. Zo wordt bij de beoordeling van verdere splitsingen van opnamegroepen daaraan geen nadere betekenis gehecht indien de floristische verschillen te gering zijn. Uitgebreide beschrijvingen van het programma TWINSPAN worden gegeven door Hill (1979) en Van Tongeren (1987).

Tegenover de voordelen van TWINSPAN, met name de snelheid van bewerking en de overzichtelijke tabel, staan ook enkele nadelen. Allereerst kan de rangschikking van opnamen en soorten langs de eerste ordinatie-as te zeer beïnvloed worden door het optreden van zeldzame soorten. Verder zijn vaak de opnamenblokken in de tabel heterogeen en onduidelijk gekarakteriseerd, omdat dezelfde soorten preferentie kunnen vertonen voor subgroepen op verschillende niveaus in de hiërarchie.

10.4.2 Het programma FLEXCLUS

Op basis van een similariteitsmatrix, waarin opnamen

met elkaar worden vergeleken, worden in agglomeratieve programma's opnamen samengevoegd tot clusters. Een voorbeeld van een dergelijk programma is FLEXCLUS, een afkorting van *FLEXible CLUStering*. Dit programma, geschreven door Van Tongeren (1986), vertoont een aantal overeenkomsten met het programma TABORD (Van der Maarel et al. 1978). Het samenvoegen van opnamen kan op verschillende manieren plaatsvinden. In het programma FLEXCLUS wordt gebruikt gemaakt van een vorm van *single linkage clustering*. *Single linkage clustering*, ook wel *nearest neighbour clustering* genoemd, houdt in dat opnamen worden samengevoegd tot kleine groepen, die op hun beurt weer aan elkaar kunnen worden gekoppeld, waarbij de afstand tussen twee groepen gedefinieerd is als de kleinste afstand tussen een opname uit de ene groep en een opname uit de andere groep. Voordat samenvoeging plaatsvindt, wordt een drempelwaarde gedefinieerd; deze waarborgt een minimale similariteit tussen elk opnamenpaar binnen een cluster (zie ook Sørensen 1948). De methode van *single linkage clustering* werkt betrekkelijk snel, maar leidt in veel gevallen niet tot een bevredigend eindresultaat omdat veelal heterogene clusters ontstaan. Om deze reden wordt in het programma FLEXCLUS de *single linkage clustering* gevolgd door een *centroid sorting procedure*, waarbij de opnamegroepen als het ware geschoond worden (zie ook Gremmen 1983). Voor elke opname wordt nagegaan of deze niet beter aan een andere groep kan worden toegevoegd. Een opname wordt verplaatst indien de afstand van die opname tot het centrum van de opnamengroep waaraan deze in eerste instantie is toebedeeld groter is dan de afstand die de opname heeft tot het centrum van een van de andere groepen (Van Tongeren 1987). Omdat na verplaatsing van een opname het centrum van zowel de groep waaruit de opname is verwijderd als van de groep waaraan de opname is toegevoegd is veranderd, kunnen na herschikking ook weer andere opnamen verplaatst moeten worden, totdat de opnamegroepen maximaal van elkaar verschillen. Het is daarom zinvol meer dan één herschikking uit te voeren.

Na de eigenlijke clustering kan het resultaat nog op velerlei manieren worden bijgesteld. Zo kunnen sterk op elkaar gelijkende groepen van opnamen samengevoegd en te heterogene groepen gesplitst worden. Ook kunnen opnamen uit een groep worden verwijderd. De volgorde van de opnamegroepen in de tabel, die wordt bepaald aan de hand van ordinatie (*reciprocal averaging*), kan interactief worden gewijzigd. Ook binnen de groepen worden de opnamen gerangschikt volgens een ordinatie. In de soortenlijst van de tabel worden constante soorten bovenaan geplaatst, gevolgd door differentiërende soorten en indifferente soorten. Al met al heeft het programma FLEXCLUS als voordeel dat goed leesbare tabellen ont-

staan. Een nadeel van het programma is dat de procedure vrij ingewikkeld is en dat de gebruiker er veel ervaring mee moet hebben om alle keuzemomenten optimaal te kunnen benutten.

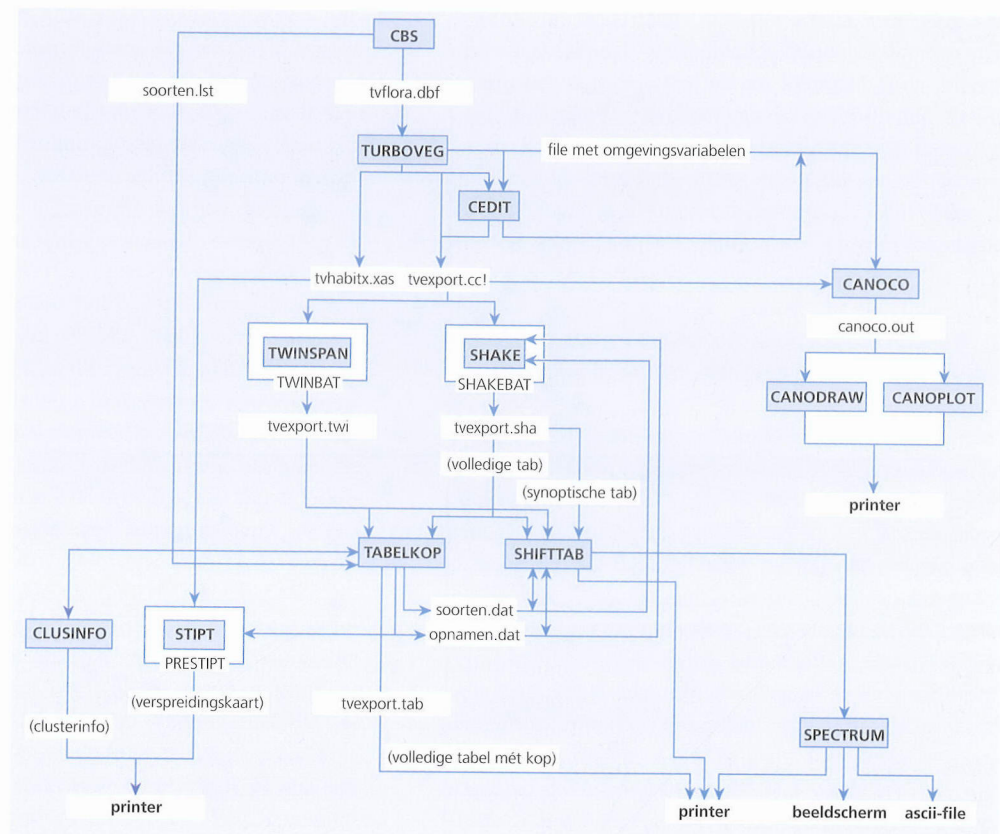
Behalve dat het programma FLEXCLUS gebruikt kan worden voor het maken van tabellen uitgaande van individuele opnamen, bestaat ook de mogelijkheid de uitkomst van TWINSPAN-bewerkingen te modificeren door de verkregen clusters voor de lagere classificatieniveaus opnieuw te bewerken. Een beschrijving van het programma FLEXCLUS wordt gegeven door Van Tongeren (1986).

10.5 Procedure van de verwerking van vegetatieopnamen bij classificatie

Om te illustreren hoe in de praktijk met programmatuur wordt gewerkt, wordt hier kort de procedure weergegeven van verwerking van vegetatieopnamen zoals deze wordt gevolgd bij de classificatie van de plantengemeenschappen van Nederland (zie ook fig. 10.2). Er wordt ingegaan op aspecten van technische aard en op enkele problemen die zich daarbij voordoen (zie ook § 6.3).

Voor de classificatie van de vegetatie van Nederland wordt gebruikt gemaakt van tienduizenden vegetatieopnamen die afkomstig zijn uit sterk uiteenlopende bronnen, zoals literatuur, persoonlijke archieven (veldboekjes) en geautomatiseerde archieven (van diverse instituten en universiteiten), en die gemaakt zijn in de periode van 1929 tot heden. Het beschikbare materiaal is derhalve nogal wisselend van aard en de wetenschappelijke bruikbaarheid loopt sterk uiteen. De opnamen worden dan ook eerst aan een kritische beschouwing onderworpen, alvorens ze al dan niet in een geautomatiseerde bestand worden ingevoerd. De criteria die hierbij zijn gehanteerd betreffen de gevolgde opnamemethodiek, de homogeniteit van de opname, de kwaliteit van de soortenlijst, en de geografische en syntaxonomische spreiding van de opnamen (zie § 6.3.1).

Het invoeren van opnamen in de computer kan op verschillende wijzen plaatsvinden. Aanvankelijk gebeurde dit door overschrijven van opnamen op standaardformulieren, die vervolgens in een computerbestand worden ingevoerd. Deze werkwijze is verlaten en heeft plaatsgemaakt voor interactief, dat wil zeggen rechtstreeks invoeren van opnamen in een computerbestand met behulp van het programma TURBOVEG (Hennekens 1994). Deze methode is veel sneller en effectiever. Het aantal fouten wordt geminimaliseerd doordat tijdens het invoe-



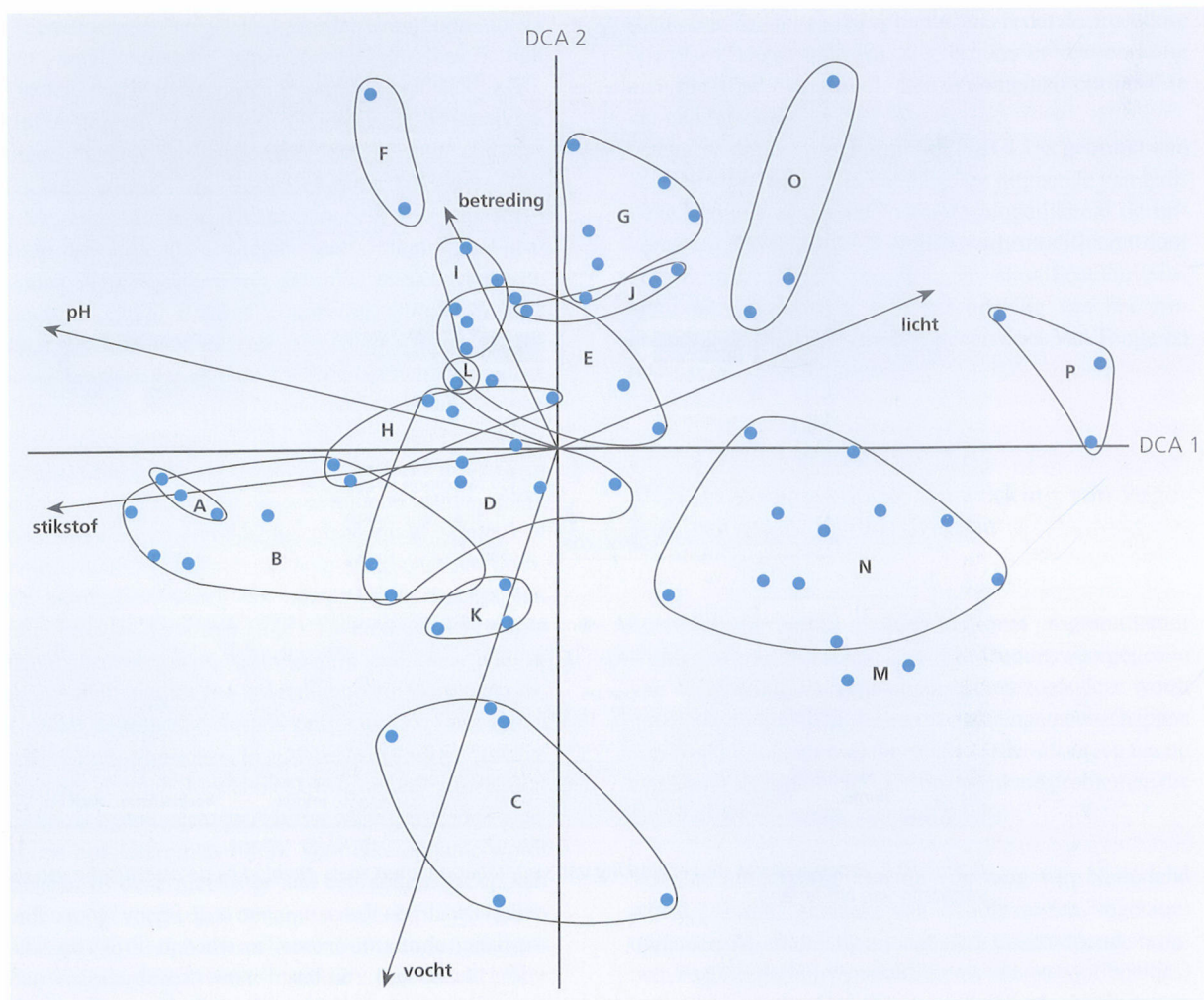
Figuur 10.2. Procedure van de verwerking van vegetatieopnamen zoals gevolgd bij de classificatie van de vegetatie van Nederland.

ren van de opnamen allerlei controles worden uitgevoerd.

Behalve de floristische gegevens worden bij iedere opname zogenaamde kopgegevens ingevoerd. Deze betreffen: opnamenummer, geografische positie (kilometerblok/atlasblok, Amersfoort-coördinaten), datum, bedekking en hoogte van boom-, struik- en kruidlaag, bedekking van de moslaag, oppervlakte, syntaxon, auteur en opnameschaal. Opnamenummer, syntaxon, auteur en opnameschaal zijn obligaat. Omdat een aantal opnameschalen van de abundantie- of bedekkingswaarden der soorten geen directe computerverwerking toelaat, worden deze omgezet naar procentuele bedekkingswaarden. De oorspronkelijke code blijft echter wel bewaard. Verder krijgt iedere opname een voorlopige syntaxoncode volgens de indeling van Westhoff & Den Held (1969). Na bewerking van een vegetatie-eenheid krijgen de opnamen een definitieve syntaxoncode volgens de nieuwe synsystematische indeling.

Kopgegevens en floristische gegevens worden bij de invoer met behulp van het programma TURBOVEG in

aparte bestanden opgeslagen, die via het unieke opnamenummer aan elkaar gerelateerd blijven. Vanuit de opnamenbestanden kunnen op velerlei wijzen selecties van opnamen plaatsvinden. Een van de belangrijkste criteria is de voorlopige syntaxonomische code. De bewerking van de Nederlandse plantengemeenschappen wordt uit praktische overwegingen namelijk per klasse aangepakt. Een bewerking van alle beschikbare opnamen (circa 180.000) op basis van alle taxa (circa 4.000) zou, afgezien van het ontbreken van voldoende reken capaciteit, als 'output' een tabel opleveren met de oppervlakte van iets meer dan een voetbalveld. Bij floristisch nauw aan elkaar verwante delen van verschillende klassen (bijvoorbeeld het *Junco-Molinion* en het *Calthion palustris* uit de klasse van de *Molinio-Arrhenatheretea*, het *Caricion nigrae* uit de *Parvocaricetea*, en het *Nardo-Galion* uit de *Nardetea*) worden de opnamen uit deze klassen ook samen bewerkt. Verder kunnen selecties gemaakt worden aan de hand van soortencombinaties, waarbij desgewenst als tweede criterium een bepaald traject van de bedekkingswaarde van de soort(en) kan gelden. Een ander veel gebruikt criterium is



- | | |
|----------|--|
| A | hoofdgroep A, eiken-beukenbossen |
| B | ruigten van stikstofrijke bodem |
| C | vochtige tot natte graslanden en ruigten van matig voedselrijke bodem |
| D | glanshaverhooilanden van relatief vochtige grond |
| E | glanshaverhooilanden van relatief droge grond |
| F | kalkgrasland en echt bitterkruid-subassociatie van de glanshaverassociatie |
| G | zandige droge graslanden en helmduinen |

- | | |
|----------------|---|
| H, I, J | tredplantengemeenschappen |
| K | zilverschoonverbond |
| L | knopbiesverbond |
| M | dopheideverbond |
| N | heide, heischraal grasland, bosbermen van schrale, zure grond |
| O | zilverhaververbond en duinviooltjesverbond |
| P | buntgras-associatie |

Figuur 10.3. Voorbeeld van een ordinatiediagram verkregen met detrended correspondence analysis (CANOCO), waarin 69 plantengemeenschappen van Nederlandse wegbermen zijn gecorreleerd met verschillende milieufactoren. De punten zijn de clustercentröiden van de onderscheiden gemeenschappen die in de groepen A t/m P zijn samengevat; de lengte van de pijlen geeft de mate van correlatie weer. De eerste as (DCA1) is gerelateerd aan de vruchtbaarheid van de bodem en de zuurgraad; de tweede as (DCA2) is vooral gerelateerd aan de vochtigheid van de standplaats (vereenvoudigd naar Šykora et al. 1993).

het jaar waarin de opname gemaakt is. Door opnamen uit verschillende perioden met elkaar te vergelijken kan een eventuele verandering in de soortensamenstelling van een syntaxon worden aangetoond.

Wat betreft het selecteren van opnamen kan nog de *two-step approach* van Van der Maarel et al. (1987) worden genoemd. Hierbij wordt het totale opnamemateriaal dat

men wil bewerken verdeeld in groepen, volgens een min of meer objectief criterium, bijvoorbeeld geografische samenhang of voorlopige syntaxonische positie (zie hierboven). Binnen elke groep worden numerieke bewerkingen uitgevoerd hetgeen leidt tot het onderscheiden van 'clusters van de eerste orde'. Alle clusters van alle groepen worden in een tweede classificatieronde met elkaar

vergeleken. Daarbij kan blijken dat clusters van verschillende groepen identiek zijn en deze worden dan in tweede instantie samengevoegd. In het voorbeeld van Van der Maarel et al. (1987), dat betrekking had op grote aantallen opnamen van vele lokale duingebieden langs de kust van de Golf van Mexico, bleek deze aanpak effectief te zijn.

Nadat een selectie is uitgevoerd, worden de opnamen omgezet in een zogenaamd *Cornell-Condensed format*. Dit formaat maakt bewerking met diverse ordinatie- en classificatieprogramma's mogelijk. Het *Cornell-Condensed*-bestand wordt in veel gevallen echter eerst nog bewerkt met behulp van CEDIT (Van Tongeren 1990) om nauw aan elkaar verwante taxa te kunnen samenvoegen tot één taxon. Wanneer bijvoorbeeld *Eleocharis palustris* s.l., *Eleocharis palustris* subsp. *palustris* en *Eleocharis palustris* subsp. *uniglumis* alle in de lijst van taxa voorkomen, worden deze desgewenst samengevoegd tot één taxon: *Eleocharis palustris* (Gewone waterbies). Vervolgens wordt met het al dan niet gewijzigde *Cornell-Condensed*-bestand een eerste TWINSPAN-bewerking uitgevoerd. Als een nadere analyse wenselijk is, wordt het bestand op basis van het TWINSPAN-resultaat gesplitst in kleinere deelbestanden, die dan opnieuw met TWINSPAN of FLEXCLUS bewerkt kunnen worden. Het verder ordenen van de verkregen tabel vindt plaats met behulp van het programma SHAKE (Van Tongeren ongepubl.), waarmee de volgorde van opnamen en soorten kan worden veranderd. Voor de interpretatie van de tabellen is het van belang dat de zogenaamde kopgegevens in de tabellen zelf af te lezen zijn. Deze toevoeging vindt plaats met behulp van het programma TABELKOP (Hennekens 1994); tevens worden met dit programma de soortscodes vervangen door volledige soortnamen. Een eventuele omzetting van plantensociologische tabellen naar synoptische tabellen vindt wederom plaats met behulp van het programma SHAKE, waarbij voor iedere soort per cluster de presentiewaarde en de karakteristieke bedekkingswaarde (de gemiddelde bedekking gedeeld door de presentie) wordt berekend.

Om de plantensociologische interpretatie van de syntaxonomische eenheden te vergemakkelijken bestaat de mogelijkheid om met behulp van het programma SPECTRUM (Hennekens 1994) ecologische en chorologische spectra te berekenen. Hiertoe worden de vegetatiekundige gegevens van de synoptische tabellen gekoppeld aan floristische gegevens, zoals die zijn opgeslagen in het Botanisch Basisregister (Centraal Bureau voor de Statistiek). Deze floristische gegevens hebben bijvoorbeeld betrekking op het uiterlijk van de soort (o.a. groeivorm), haar relatie tot de abiotische omstandigheden (o.a. grondwaterafhankelijkheid, levensvorm, Ellenberg-waarden) en plantengeografische aspecten (o.a. zeldzaamheid,

ligging van Nederland t.o.v. het areaal, Europese verspreiding). De analyse resulteert in frequentiediagrammen, waarin per vegetatietype en per kenmerk (bijv. levensvorm) het relatieve aandeel van de verschillende klassen van dat kenmerk (bijv. therofyt, hemicryptofyt, fanerofyt, enz.) van de soorten die in dat vegetatietype voorkomen wordt gegeven (Van Duuren & Schaminée 1990). Meestal worden zogenaamde kwantitatieve spectra berekend, waarbij de berekening plaatsvindt op basis van de presentiewaarden der soorten in de synoptische tabellen; dit in tegenstelling tot kwalitatieve spectra, waarbij slechts wordt uitgegaan van het al dan niet aanwezig zijn van soorten.

Voor het analyseren van de betrekkingen tussen de onderscheiden syntaxa en afzonderlijke milieufactoren kan het programma CANOCO (Ter Braak 1988) worden toegepast, een techniek die voortbouwt op het ordinatieprogramma van het DECORANA-pakket. Uiteraard dienen hiervoor de milieufactoren bekend, dat wil zeggen gemeten te zijn. Door middel van een groot aantal assen worden vele factoren in één figuur samengevat. De plaats van ieder syntaxon in het ordinatiediagram (weergegeven als één punt of door een groep van punten die de opnamen vertegenwoordigen) wordt bepaald door de scores op de afzonderlijke assen (zie fig. 10.3). Ook de afzonderlijke soorten kunnen met behulp van CANOCO worden uitgezet tegen de milieufactoren.

10.6 Enkele kanttekeningen bij het gebruik van numerieke methoden

Voor de verwerking van grote aantallen vegetatieopnamen zijn numerieke classificatiemethoden wenselijk, zo niet noodzakelijk. De vragen hierbij zijn hoe de computer optimaal gebruikt en aan de hand van welke criteria het resultaat van numerieke bewerking beoordeeld kan worden. Eén van de criteria voor de beoordeling is dat het verkregen classificatiesysteem bruikbaar moet zijn, dat wil zeggen dat het ecologisch goed interpreteerbaar is, hetgeen mede door de onderzoeker bepaald wordt. De kwestie is in hoeverre de gevolgde procedure dan nog objectief is, respectievelijk objectief zou moeten zijn.

Alleen al het feit echter dat bij het gebruik van computer-methoden uit zeer veel verschillende opties of combinaties van opties kan worden gekozen, maakt dat het eindresultaat net zo min objectief genoemd kan worden als het resultaat dat wordt verkregen via de handmatige methode.

Niet alleen de verwerking van de gegevens, maar ook de manier waarop deze verzameld zijn is van invloed op het

uiteindelijke resultaat. Zo hangen de keuze van de afbakening en de grootte van het proefvlak (onder meer bepalend voor de homogeniteit), en vooral ook de beslissing waar de opname gemaakt wordt, sterk af van de ervaring en het persoonlijke oordeel van de onderzoeker en deze zijn dus in zekere mate subjectief. Met betrekking tot het gebruik van opnamen uit de literatuur dient erop gewezen te worden dat verschillende onderzoekers ieder min of meer hun eigen methode volgen en hun eigen (al of niet systematische) 'fouten' maken; bovendien kunnen in de loop van de tijd taxonomische opvattingen zijn veranderd. Een probleem bij de beoordeling van het te gebruiken opname materiaal is bovendien de geografische verspreiding van de opnamen en de verdeling over de syntaxonomische eenheden.

Een wezenlijk probleem bij de numerieke verwerking van opnamen is dat uitsluitend de floristische samenstelling (al of niet met inbegrip van de bedekkingswaarden) in de berekeningen wordt betrokken, terwijl bij de classificatie volgens de Frans-Zwitserse school tot op zekere hoogte ook met andere aspecten rekening wordt gehouden, zoals met de vegetatiestructuur. Bijstelling van de resultaten 'met de hand' kan om die reden nodig zijn.

Een ander probleem hangt samen met het verschil in de ecologische informatiewaarde van soorten. Volgens Barkman zou voor een objectieve benadering bij classificatie aan iedere soort een gewicht toegekend moeten worden dat overeenkomt met zijn ecologische informatiewaarde (die dan alleen regionaal geldig is). Hoewel hiertoe wel aanzetten zijn gegeven (Noest et al. 1989; zie ook § 10.2), is dit vooralsnog moeilijk voor alle soorten te realiseren; bovendien blijft het de vraag in hoeverre een dergelijke weging objectief is.

Naast floristische verschillen (in presentie en/of bedekking) kunnen ook verschillen in vitaliteit of fertiliteit van soorten een rol spelen bij het onderscheiden van vegetatie-eenheden. Planten van *Orchis purpurea* bijvoorbeeld kunnen bloeiend dan wel vegetatief voorkomen, hetgeen samenhangt met een verschil in beschaduwing. Hetzelfde geldt voor bloeiende exemplaren van *Nymphaea alba*

(Witte waterlelie), die een andere diagnostische betekenis hebben dan niet-bloeiende; de laatste bevinden zich gewoonlijk in een later stadium van de successie. In de beschikbare computerprogramma's kan deze informatie niet of slechts langs een moeizame weg rekenkundig verwerkt worden.

Volgens welke criteria opnamegroepen van elkaar dienen te worden gescheiden en wat de syntaxonomische status is van het onderscheid (niveau van klasse, orde, verbond, associatie of subassociatie) hangt onder andere af van het totale aantal soorten in de opnamen. Een verschil van twee soorten tussen opnamen van *Lemna*-begroeiingen kan aanleiding zijn tot het onderscheiden van twee verbonden, terwijl een verschil van twee soorten tussen opnamen van voedselrijke hellingbossen nog geen aanleiding hoeft te zijn tot het onderscheiden van twee subassociaties.

Het oordeel over de status van vooral de hogere syntaxonomische eenheden kan niet uitsluitend plaatsvinden op basis van Nederlands opnamemateriaal. Hiervoor zullen ook verwante vegetatietypen in de ons omringende landen bestudeerd moeten worden, bijvoorbeeld aan de hand van literatuur. Een voorbeeld waarbij buitenlandse gegevens en inzichten doorslaggevend zijn, betreft de classificatie van de Nederlandse muurbegroeiingen, die in ons land slechts fragmentair ontwikkeld zijn.

Uit het bovenstaande moge blijken dat een kritische betrokkenheid van de onderzoeker dus voortdurend gewenst blijft. Dit kan leiden tot bijsturing, waarmee invloed wordt uitgeoefend op het uiteindelijke classificatieresultaat. Deze opvatting komt dus niet overeen met het idee dat de computer allesbepalend is en als een soort trechter gezien kan worden waarin men alle opnamen stopt, waarna de vegetatie-eenheden er keurig geïnclassificeerd uitrollen; de computer zou hierin als een soort *black-box* fungeren, waarbinnen het wonder gebeurt. Men spreekt in dit verband van het autoriteitssyndroom: de misvatting dat de technieken objectief zijn en het clusterresultaat als definitief geaccepteerd moet worden.